AIR FORCE RESEARCH LABORATORY TEST AND EVALUATION, VERIFICATION AND VALIDATION OF AUTONOMOUS SYSTEMS

CHALLENGE EXPLORATION FINAL REPORT

DISTRIBUTION STATEMENT A. Approved for Public Release, Distribution Unlimited Case Number 88ABW-2014-4063
Other requests shall be referred to AFRL/RQQA, 2210 8th Street , Wright-Patterson AFB, OH 45433

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
13 NOV 2014	**Final**	**23 MAR 2013 - 07 JUL 2014**

4. TITLE AND SUBTITLE
AIR FORCE RESEARCH LABORATORY TEST AND EVALUATION, VERIFICATION AND VALIDATION OF AUTONOMOUS SYSTEMS CHALLENGE EXPLORATION FINAL REPORT

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Matthew Clark, Kris Kearns, Dr. Jim Overholt, Kerianne Gross, Bart Barthelemy, Cheryl Reed

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Control Automation Branch AFRL/RQQA 2210 Eighth St, Room 301 Wright-Patterson AFB, OH 45433

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/RQQA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT
Verification and Validation techniques will need to be adapted to address the unique challenges of developing autonomous systems. Looking towards the future, the question must be asked, how do we supplement traditional test and evaluation methods? In addition to new technology, what T&E workforce education requirements need to be addressed? As highly autonomous systems become more of a reality, trust in the system operation will transfer from the human operator to highly complex software and systems. Quantifying that trust and then providing a certification argument is a daunting task and an enabling technology for the transition of the next generation of autonomous systems. This report documents the year long study to identify, understand and categorize the unique challenges to the certification of safety critical autonomous systems by identifying the Verification and Validation (V&V) approaches needed to overcome them. The desired end state is to have a document categorizing the challenges into 3-6 thrust areas with a semi-detailed list of complementary approaches to address each challenge. The outcome of this study supports the AFRL Autonomy S&T strategy as well as provides input to the DoD Autonomy TEV&V Portfolio.

15. SUBJECT TERMS
Verification, Validation, Test & Evaluation, Trust, Autonomous Systems, Autonomy, Safety, Security, Certification

<table>
<tr><td colspan="3">16. SECURITY CLASSIFICATION OF:</td><td rowspan="2">17. LIMITATION OF ABSTRACT
UU</td><td rowspan="2">18. NUMBER OF PAGES
31</td><td rowspan="2">19a. NAME OF RESPONSIBLE PERSON</td></tr>
<tr><td>a. REPORT
unclassified</td><td>b. ABSTRACT
unclassified</td><td>c. THIS PAGE
unclassified</td></tr>
</table>

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Table of Contents

NOTE: This report references several supporting appendices. These appendices are marked Distribution C and are available upon request to government and government contractors only.

T&E, V&V of Autonomous Systems Background

Rapid development and application of new, emerging technologies in recent years are enabling autonomous systems to do amazing things in adapting to an unpredictable world – unpredictable due to system faults and failures, human error, developing weather patterns, air turbulence, road conditions, changes to mission objectives, adversarial environments, etc. In applications where human life is at risk, such as defense, transportation, nuclear, and medical applications, establishing justifiable confidence that these self-governing systems are safe poses one of the greatest challenges to large-scale acceptance of autonomous systems.

Currently, the burden of providing certification authorities assurances of safety-critical systems rests on the test and evaluation of such systems.[1] As autonomous systems become more complex, the notion that systems can be fully tested is becoming more difficult, especially as higher levels of self-governing systems become a reality. As these systems react to more environmental stimulus and have larger decision spaces, testing all possible states and all ranges of inputs to the system becomes an unachievable goal. As of 2012, autonomous cars had completed over 300,000 miles of testing without incident; are they safe for the general public?[2] Were all the requirements of safe driving captured? How much of the software was actually exercised? How many of the inputs were covered and were all interdependencies covered? How well do the test conditions match real-world conditions? What unknown system behaviors still exist?

Verification and Validation techniques will need to be adapted to address the unique challenges of developing autonomous systems. This change was highlighted in the 2010 Air Force Technology Horizon report, which stated that, ``It is possible to develop systems having high levels of autonomy, but it is the lack of suitable V&V methods that prevents all but relatively low levels of autonomy from being certified for use."[3]

Looking towards the future, the question must be asked, how do we supplement traditional test and evaluation methods?[4] In addition to new technology, what T&E workforce education requirements need to be addressed?[5]As highly autonomous systems become more of a reality, trust in the system operation will transfer from the human operator to highly complex software and systems. Quantifying that trust and then providing a certification argument is a daunting task and an enabling technology for the transition of the next generation of autonomous systems.

Research Exploration Objective

The goal of this AFRL activity, facilitated by the Wright Brothers Research Institute, was to identify, understand and categorize the unique challenges to the certification of safety critical autonomous systems by identifying the Verification and Validation (V&V) approaches needed to overcome them. The desired end state is to have a document categorizing the challenges into 3-6 thrust areas with a semi-detailed list of complementary approaches to address each challenge. The outcome of this study supports the AFRL Autonomy S&T strategy as well as provides input to the DoD Autonomy TEV&V Portfolio.

---

[1] Kelly J, Hayhurst, et al. "A Practical Tutorial on Modified Condition/Decision Coverage." (2001).
[2] F. Lardinois. Google's self-driving cars complete 300k miles without accident, Aug 2012. http://techcrunch.com/2012/08/07/google-cars-300000-miles-without-accident/
[3] Dahm, W. J. A. "Technology Horizons a Vision for Air Force Science & Technology During 2010-2030." USAF HQ (2010).
[4] Fisher, Michael, Louise Dennis, and Matt Webster. "Verifying autonomous systems." Communications of the ACM 56.9 (2013): 84-93.
[5] Davis, Jennifer A., et al. "Study on the Barriers to the Industrial Adoption of Formal Methods." Formal Methods for Industrial Critical Systems. Springer Berlin Heidelberg, 2013. 63-77.

Executive Summary

Today

The current state of Test and Evaluation for Autonomous systems is through exhaustive modeling and simulation and testing. Current M&S or T&E methods, though effective for countless currently fielded systems, become a bottleneck when attempting to field higher levels of Autonomy. Following the three workshops, the meeting owners met several times to refine the technical challenges and produce the initial draft of the AFRL TEV&V storyboard and supporting research needs. This chapter highlights the key components of the storyboard. The graphical representation can be found in Appendix A11.

Limitations to the Current Certification Process

Five limitations to the current certification process were consolidated from the information provided in the three workshops.

- Test for all known conditions
- V&V is late in design process
- Difficult to objectively measure risk
- Decision making burden on humans
- System level test for small changes

Enduring Problems

The team found four enduring problems:

- State-Space Explosion – Autonomous cognitive agents are by definition learning / adaptive in nature. The algorithmic decision space is non-deterministic, i.e. the output cannot be predicted due to multiple possible outcomes for each input. This space cannot be exhaustively searched, examined, or tested. The model under test exponentially explodes the more it is refined to adequately test all known conditions, factors, interactions.

- Unpredictable Environments – The power of autonomous agents is the ability to perform in unknown, untested environmental conditions. Currently fielded systems have a very limited robustness to dynamic / changing environmental conditions. Adaptive / Autonomous algorithms have the potential capability to overcome current automated system brittleness in future dynamic, complex, and/or contested environments. However, this performance increase comes with the price of assuring correct behavior is a countless number of environmental conditions. This problem exacerbates the state-space explosion problem.

- Emergent Behavior – Interactions between systems and system factors may induce unintended consequences. With non-deterministic, adaptive systems, how do you capture all interactions between systems sufficiently to understand all intended and unintended consequences? What limitations are there with the current Design of Experiments approach to test vector generation when considering adaptive decision making in both discrete decision logic and continuous variables in an unpredictable environment?

- Human-Machine Communication –Handoff, communication, and interplay between operator and autonomy becomes a critical component to the trust and effectiveness of an autonomous system. Current certification processes eliminate the need for “trust” through exhaustive M&S and T&E to exercise all possible operational vignettes. When this is not possible at design time, how will we ensure trust in the system, what factors need to be addressed, how do we define the transparency and communication requirements for the autonomy?

Future

The future state envisions an “autonomous agent” no longer or restricted by the inability to be certified as trustworthy at an acceptable level of operation and risk. The “autonomous agent” can take many forms or provide the reasoning decision maker for manned / unmanned aircraft, cyber agents, satellites, and weapons.

Vision

The vision depicts a future state were **alternate evidence** of verification and validation can be generated through methods in addition to M&S and T&E. The results from these methods can be recorded in a **modular** fashion, enabling **compositional** verification of autonomous subcomponents at appropriate levels of abstraction, thereby reducing the system level V&V challenge. Additionally, similar to case law, well defined and iteratively developed autonomous agents will be able to establish **precedence** through past performance and “training” as a method of certification. Finally, development of autonomous agents will be **iterative, continuous**, and evolutionary, reducing the software development cycle burden.

Technical Goals to Achieve the Future State

The following technical goals provide multiple, additive methods to Verify and Validate Autonomous Systems.

Cumulative Evidence through RDT&E, DT & OT

- ***Progressive sequential modeling, simulation, test and evaluation***

Currently, Modeling and Simulation, Test and Evaluation at each Technical Readiness Level provide an invaluable resource not only to verify and validate that a system satisfies the user requirements but also to aid in technology development and maturation. However, effective methods to record, aggregate, and reuse T&E results remain an elusive and technically challenging problem. Through progressive sequential modeling, simulation, test and evaluation, how can the results from experiments performed in research and development help reduce the factor space in final operational tests? Can early experimental results be encoded based on operational test conditions, assumptions and then parsed within a database, leveraging the results to reduce the testing burden?

Statistics-based design of experiments methods currently lacks the mathematical constructs capable of designing optimized test matrices for non-deterministic software. Software systems require a risk-mitigation methodology offering the same spirit as DOE but using a Non-Statistical approach; there are sciences, methods, and tools that have utility, though they lack a codified set of overarching principles.

Evidence generated during design

- ***Guarantee appropriate decisions with traceable evidence***

Through design for certification (formalized design) at the beginning, substantial gains can be realized throughout the development and sustainment lifecycle. In order to provide assurance for machine intelligence and decision-making in complex, uncertain, and dynamic environments a paradigm shift must be realized. Similar to the early development of control theory, formal methods and analysis seeks to provide proofs about the safety, reliability, and robustness of software systems. How can verification and validation artifacts be embedded in “correct by construction” design to reduce the test and evaluation burden?

Requirements Development and Analysis

- ***Precise, structured standards to automate requirement evaluation for testability, traceability, and de-confliction***

To maximize the operational gains of advanced autonomy, design for certification must be accomplished in early requirements development. Using formalizable, mathematically rigorous natural language to specify requirements forces the subject matter expert to be explicit, clearly defining assumptions. Additionally, these formalisms provide high level operational assumptions and interoperability guarantees that can be analyzed early in the design phase. Finally, a requirement isn't complete without understanding how you will test it. Formalized requirements enable automatic test generation and traceability to low level designs.

Decision Assurance

- ***Real time monitoring and migration of undesired decisions and behaviors***

For the most demanding, most adaptive (and non-deterministic) problems, we may need even a more dramatic shift. Currently, we attempt to prove systems correct via verification of every possible state PRIOR to fielding the system. However, if, through the use of run-time architecture, we can provably bound systems behavior, then it may be possible to reduce the reliance on comprehensive off-line verification, shifting the analysis/test burden to the more deterministic run-time assurance mechanism. Provable performance bounds must be formulated to reduce the reliance on comprehensive, off-line verification, shifting more of the analysis and testing burden to more provable run-time assurance technologies. Safe operation of an autonomous system must be ensured even though the machine's behavior/performance may not be exhaustively verified according to current development or certification standards. Key tenets are to reduce the amount of testing through up-front analysis and to reduce the burden for off-line certification through run-time assurances.

Compositional Case Generation

- ***Enable reusable evidence building blocks***

All other goals focus on the design or testing of an Autonomous system. The assumption is that no one method for verification and validation will be adequate for the complexity presented by these systems. Therefore, not only do multiple methods need to be employed throughout the lifecycle, a new research area needs to be investigated in formally verifying that the composition of evidence is valid. An Assurance (or more commonly Safety) case can be defined as a structured argument, supported by evidence, intended to justify that a system is acceptably safe and secure; required as part of a regulatory process, a certificate of safety being granted only when the regulator is satisfied by the argument presented. Research must be done to formalize safety cases for the purposes of analysis and reuse. New V&V methods must eliminate excessive certification as heterogeneous machines are combined into systems during fielding. Preventing unintended emergent behavior as systems are composed into System of Systems will allow systems to be evaluated at the individual machine level while maintaining safety guarantees at the system level. Also, this technology must allow one element of a fractionated capability to be modified while minimizing the re-certification requirements of other components. The effort concentrates on reducing the reliance on the sum of the individual certifications and increase reliance on a system of systems wide certification that accounts for unintended / undesired emergent behavior.

Industry Exploration Workshop - October 23-24th 2013

This section attempts to capture the discussion and outcome of the first of three workshops facilitated by the Wright Brothers Research Institute. This workshop included key individuals from Industry that have experience with creating and fielding autonomous systems. The Wright Brothers Institute did not limit the participants to just DOD nor to just AF. Additionally, the invitation was not intended to exclude other applications of autonomous systems not in the traditional robotic or vehicle space. Autonomous cyber systems and medical systems were also welcome. The goal was to have a wide swath of the autonomous work in industry represented. That being said, most of the participants were DOD contractors and almost all of them were concentrating on the Autonomous Vehicle domain. The participants and their contact information are listed in the table below:

Wright Brothers Institute Facilitators

Cheryl Reed	WBI
Bart Barthelemy	WBI

AFRL Sponsors

Kerianne Gross	AFRL/RQQD
Kris Kearns	AFRL/RH
Jim Overholt	AFRL/RH
Matthew Clark	AFRL/RQQA

Industrial Participants

Todd Belote	Lockheed Martin
Siddhartha Bhattacharyya	Rockwell Collins
Kevin Donaghy	Lockheed Martin
Steve Dues	-
Scott Grigsby	Ball
Andreas Hofmann	Vecna
Jeff Hughes	Tenet3
Todd Jackson	Draper
Troy Jones	Draper
Adam MacDonald	Avinc
David Musliner	SIFT
Michael Niestroy	Lockheed Martin
Russ Purtell	Northrop Grumman
Tim Quellhorst	Crown
Jim Schloemer	Crown
Greg Tallant	Lockheed Martin
Thomas Weaver	Boeing
Ron Ziegler	Crown
Andrew Zimdars	Lockheed Martin
Dan Zwillinger	Raytheon
Anne Selwyn	Raytheon
Curtis Wray	Ball
George Rodgers	Northrop Grumman

Table 1: Industry Participants

Workshop Agenda

1. *DAY 1 - PROBLEM SPACE EXPLORATION*
    1.1. *ALIGN* – Introductions and Objectives
        1.1.1. DOD Autonomy Intro – Kris Kearns, AFRL
        1.1.2. TEV&V for AFRL – Matt Clark, AFRL
        1.1.3. How do we do it today? - Dr. Darryl Ahner, Air Force Institute of Technology, Director of the DoD Test and Evaluation Center of Excellence
    1.2. *EXPLORE-PARTICIPANT INTROS – Role in TEV&V, Biggest tech challenge*
    1.3. *OPEN APERTURE – Discussions*
        1.3.1. What does certified mean?
        1.3.2. What is an autonomous system?
        1.3.3. What is certification of an autonomous system?
    1.4. *CONCEPT MAP DEVELOPMENT*

2. *DAY 2– CONVERGE ON CHALLENGES*
    2.1. *ORIENT - Revisit Concept Map for new insights*
    2.2. *DISTILL KEY ELEMENTS*
        2.2.1. *Identify major technical thrusts from Concept Map*
        2.2.2. *Elaborate on/refine each technical thrust – what needs to be done to accomplish each*
    2.3. *SEQUENCE CHALLENGES AND DEPENDENCIES*
        2.3.1. *Determine any required sequencing of the various technical thrusts*
    2.4. *WRAP UP*
        2.4.1. *What is your biggest takeaway?*
        2.4.2. *What would you ask / tell the Academic Workshop Participants?*

Workshop Notes

1. DAY 1 - PROBLEM SPACE EXPLORATION

1.1. *ALIGN* – Introductions and Objectives

1.1.1. DOD Autonomy Intro – Kris Kearns, AFRL

The event started with an introduction to the DoD Autonomy Technology Challenge Areas: Human and Machine Teaming, Scalable Teaming, Machine Reasoning and Intelligence, and ***Test and Evaluation, Verification and Validation***. Then the AFRL Autonomy Goals were identified: Deliver flexible autonomy systems with highly effective human-machine teaming, Create actively coordinated teams of multiple machines to achieve mission goals, Ensure operations in complex, contested environments, and ***Ensure safe and effective systems in unanticipated and dynamic environments***.

It was highlighted that the DOD and AFRL goals, although not the same, are similar. Specifically, the DoD Autonomy TEV&V Portfolio goal is directly attributed to the AFRL Goal of "ensuring safe and effective systems." The introduction demonstrated the AF and DOD need to understand and invest in new methods of assuring the safety and operability of up-and-coming autonomous systems. Additionally, Ms. Kearns is responsible for coordinating an AFRL strategy for Autonomy. An overarching AFRL strategy document was signed as of Friday, Oct 25th. The strategy now needs "meat on the bones" highlighting a portfolio that will accomplish the strategy goals. The human-machine teaming portfolio was created first and the TEV&V was determined to be the next portfolio to be tackled. This workshop is the first of three to aid in that goal.

1.1.2. TEV&V for AFRL – Matt Clark, AFRL

Mr. Clark continued the conversation by highlighting how the information from the next three workshops would be used to provide the portfolio needed to accomplish the AFRL autonomy strategy. First, the industrial based workshop would identify the key certification challenges facing some of the DoD and other industries trying to enable further autonomous systems. The plan is to take the outcome of the meeting this week to feed to a similar academic workshop. The compiled information would then be distilled and coordinated with a broader AFRL Autonomy team. This team would consolidate and make recommendations for future TEV&V investments. Additionally, in parallel, the DoD Autonomy TEV&V initiative is working to put together similar strategies. The inputs from the above three workshops will feed the greater DoD Autonomy TEV&V effort, highlighting the investment opportunities and research needs.

1.1.3. How do we do it today? - Dr. Darryl Ahner, Air Force Institute of Technology, Director of the DoD Test and Evaluation Center of Excellence

Dr. Darryl Ahner then presented the "state of the art" in testing for software intensive systems, highlighting the fact that there is a continuum of methods currently used depending on the software risk type and measure. He highlighted the main areas of software risk as Stochastic, Deterministic, and Probabilistic. The main risk of future autonomous systems is their non-deterministic / learning behavior. How do guarantee measure software coverage when the system learns while being tested? Dr. Ahner's slides are attached in appendix A1.

1.2. EXPLORE-PARTICIPANT INTROS – Role in TEV&V Biggest tech challenge

During the section each participant was asked to provide some information about their background and one biggest challenge they face in TEV&V right now. These challenges were captured but not used until the end of the first day for the purposes of comparing the initial thoughts with the group challenges. Some of the challenges included:

- Validating interaction of Autonomous System in a hostile environment
- How to produce a measure of effectiveness
- How to protect from emergent behavior
- How to make better , more trustable man-machine interfaces
- Requirements validation -> metrics / types of requirements for autonomy
- Acceptance of stochastic processes
- Once certified, will end users trust to use them?
- Deployment of academic research to industry

- Hand off of control between human and machine
- Writing formally what you want (software)
- Do we have appropriate use cases?
- Better standards
- Prediction of what you expect to happen

1.3. OPEN APERTURE – Discussions

The next activity involved group discussions at each of the six tables asking three questions. The purpose of the exercise was to highlight the differences in how we define certification, autonomous, and certification of autonomous systems.

1.3.1. What does certified mean?

Some of the responses for what does certified mean:

- Oversight authority has determined that items have specific standards
- Meets a standard repeated / reliably determined by experts
- Can be trusted to perform under certain conditions
- Tested to meet a standard (test, lab experiment, historical data) based from specific use cases

Several common threads emerged:

- There is an authority that sets some standards for safety / reliability
- This authority has oversight into what requirements are levied on systems to comply with the standards established
- Verification and Validation methods (predominately test) show compliance to standards

An interesting discussion emerged between the DoD and the non-DoD (manufacturing) industry representatives about liability. Crown, a forklift manufacturing company, observed that risk was assumed by the plant manager and ultimately liability for accidents. The plant manager used the standards as compliance guidelines but was ultimately responsible for updating and maintaining compliance. However, they also observed liability was not treated the same for the aerospace industry.

1.3.2. What is an autonomous system?

The discussion about what an autonomous system is ranged wildly (as expected). The discussion culminated on the "levels" of autonomy highlighted by several communities.

Dr. Overholt and Ms. Kearns addressed the "levels" of autonomy as stated in the Defense Scientific Board Report on Role of Autonomy in DoD systems quoting, "The Task Force recommends that the DoD abandon the use of 'levels of autonomy' and replace them with an autonomous systems reference framework that embraces three-facets; cognitive echelon, mission timelines, human-machine system trade spaces."[6] For the follow-on efforts, Ms. Kearns brought out the AFRL definition of Autonomy to guide the discussion:

"Automation: The system functions with no/little human operator involvement; however, the system performance is limited to the specific actions it has been designed to do.[7] Typically these are well-defined tasks that have predetermined responses, i.e. simple rule-based responses.

Autonomy: Systems which have a set of intelligence-based capabilities that allow it to respond to situations that were not pre-programmed or anticipated in the design (i.e. decision-based responses). Autonomous systems have a degree of self-government and self-directed behavior (with the human's proxy for decisions)."[8]

---

[6] DoD Defense Science Board Report: Role of Autonomy in DoD systems, Dr. Paul Kaminski (DSB Chair), July 2012
[7] Spacecraft Autonomy Technology: A Survey. Erwin, R. Scott and Paul Zetocha, AFRL/RV
[8] Air Force Research Laboratory Autonomy Science and Technology Strategy: Maj Gen Masiello, Oct 2013

1.3.3. What is certification of an autonomous system?

Definitions on what it means to certify an autonomous system relegated to combining the similarities between the answers to the two above questions. Ultimately resulting in a definition similar to the following: an authority sets standards for safety / reliability for a system that responds to situations that were not pre-programmed or anticipated in design.

1.4. CONCEPT MAP DEVELOPMENT

The second half of the first day focused on taking the assumptions about certification and autonomy and using them to produce a concept map. Each table was directed to come up with the 5 top challenges to the TEV&V of autonomous systems independently. Once completed, each table was directed to consolidate "like" or common challenges. Finally, each table took turns placing their consolidated challenges up on the white board, again grouping them based on similarity. The concept map from the first day contained the following challenges listed below. The entire concept map can be accessed in appendix A2.

Culture Change - Make TEV&V research more attractive, agile, promotes evolution
Trust - Establishing Acceptable Risk
Human/Machine Interaction - Methods for Mixed Human Intelligent Machines
Requirements - Have Appropriate / Formal / Accepted Requirements for Autonomous Systems
Systems of Systems V&V-Validating and Composing Interactions of Autonomous Systems
Synthesis - Correct By Construct Synthesis of Systems from Design
Uncertainty - Formal Representation and Characterization of Uncertainty
Emergent Behavior- Resolve the Paradox of the Desire for Novel Behavior and the Req for No Bad Behavior
Test - Create Acceptable/Sufficient Test Success
Runtime Verification - On Board Verification, Runtime Safety Monitoring
Security - Security of Autonomous Systems and Their Use
Modeling and Simulation - Establishing an Autonomous System Virtual Proving Ground
Tool Verification - Formal Verification Tools, Proving Tools Perform as Expected

Finally, Policy was highlighted as a "parking lot" challenge for the purposes of isolating the technical barriers from the social / economic ones. These challenges received their own grouping.

Parking Lot Challenges - Societal Acceptance of Systems That Fail, User/Society Trust in Autonomous Systems

2. *DAY 2– CONVERGE ON CHALLENGES*

2.1. ORIENT - Revisit Concept Map for new insights

An interesting thought emerged at the beginning of the second day. The group highlighted the concept of “Training” and then “Licensing” a human vs. Certifying a machine or system. The question was asked, “If the certification paradigm changed to a training and licensing model for autonomy, how would verification and validation change for these systems?” This resonated deeply with the entire group, which caused us to create an additional category on the concept map; “Defacto Standards of Safety Based on a Licensing vs. Certification Paradigm.”

2.2. DISTILL KEY ELEMENTS

2.2.1. Identify major technical thrusts from Concept map

This exercise consolidated earlier brainstorming into major technical thrusts or goals needed to achieve the overarching verification and validation of autonomous systems goal. The technical thrusts are identified and described in section 2.2.2.

2.2.2. Elaborate on/refine each technical thrust – what needs to be done to accomplish each

The next exercise took the bulk of the morning and into the afternoon. The participants were directed to take each top level challenge identified previously and perform a mini Goals, Objectives, Technical Challenges, and Approaches (GOTCHA) exercise. The top level challenges were considered the Goal. Each team then identified the objective and technical challenges to be addressed to achieve the goal. The revised concept map with Objectives and Technical Challenges can be found in appendix A3.

Trust

Objective:

• What suite of tests and what type of evidence leads to trust or acceptance in Autonomous Systems?

Human/Machine Interaction

Objective:

• How to test effectiveness of human/machine team?

Requirements

Objective:

• Develop an industry standard for developing system level requirements.
• Requirements should be: Testable, Precise (Formal)

System of Systems

Objective:

• To ensure safe and effective operation of a system through composing certified/licensed components/subsystems/systems

Synthesis

Objective:

• To automatically create an executable autonomous system from models of environment, capabilities, requirements, constraints, and goals.

Uncertainty

Objective:

• Create determined TEV&V methods for (non)-deterministic systems in non-deterministic environments
• Objective doesn’t capture all of uncertainty, can we determine when T&E 70% solution is “good enough?”

Emergent Behavior

Objective:

• Mitigate negative effects of Emergent Behavior
• Assess and manage the effects of Emergent Behavior within the TEV&V process

Test

Objective:

• Effective evaluation of autonomous behavior in a cost-effective manner

Runtime Verification

Objective:

• Provide fulltime active monitoring for unexpected behavior and/or events and provide appropriate supervisory control.

Security

Objective:

• Protect Autonomous Systems from unauthorized control/access

Modeling and Simulation

Objective:

• Provide cost and time effective means to develop, T&E, V&V Autonomous Systems

Tool Verification

Objective:

• Develop new verification tools for non-deterministic Autonomous Systems

Defacto Standards Licensing vs Certification

Objective:

• An assessment to determine if a system should be approved as having reached a certain level of reasonableness for competence/ safety within its intended operating environment.

2.3. SEQUENCE CHALLENGES AND DEPENDENCIES

2.3.1. Determine any required sequencing of the various technical thrusts

The next ½ of the day focused on prioritizing the Goals and Technical Challenges

Section	Rating	Red	Yellow	Green
System of Systems V&V	B	3	8	7
Uncertainty	?	6	6	5
Emegent Behavior	C	1	3	18
Defacto Standards Licensing vs Certification	A	14	2	1
Requirements	A	14	5	11
Modeling and Simulation	A	15	2	2
Trust	B	4	10	3
Human/Machine Interaction	B	2	12	4
Test	A	11	6	
Runtime Verification	B	2	11	5
Synthesis	C	1	4	15
Tool Verification	B	3	9	4
Security	?	4	9	8

Figure 2: Priority and Score for Each Technical Challenge

(Red (A) = Near Term, Orange (B) = Mid Term, Green (C) = Far Term)

WRAP UP

2.3.2. What is your biggest takeaway?

The following list highlighted the biggest takeaways

- Licensing Paradigm was by far the biggest takeaway and AHA moment. If technology could enable a licensing of autonomous algorithms combined with a certified system great gains could be made in this area. 9 people highlighted this as the number 1 takeaway!

Other takeaways:

- Why isn't there more investment in this area if this is such a hard problem?
- Requirements generation and validation is huge for complex systems let alone autonomous systems!
- More formalism is needed
- This is a hard problem that will take a lot of time, start small
- There is a public perception that we (Defense) is farther along in autonomy and certification. A response was made that we are pretty far along in autonomous systems; it's just that to go any further with the technology a new certification paradigm has to be in place. "We have hit a brick wall"
- Can requirements be written for autonomy without saying the word "autonomy?" What is wrong with how we write requirements?
- Cultural limitations are going to precede technical ones
- This problem can only be solved when a platform is defined

2.3.3. What would you ask / tell the Academic Workshop Participants?

- Strong Formal Methods presence at the workshop, non-deterministic aspect to FM
- Definition of autonomy; machine learning for V&V
- Formal Methods for uncertainty – stochastic testing alternative
- Ask them which technology challenge they would address first
- Huge payback on synthesis of requirements to designs – how do we do this?
- Help us do TEV&V better with more formalization
- Formal methods for Runtime Verification and uncertainty
- Roadblocks for stochastic test methods
- How do we figure out the human-machine interaction?
- How viable and what else is there other than stochastic testing?
- Figure of merit (non-determinism) how do we come up with standardized levels of goodness or risk?
- What is the difference between certification w/ man as the pilot vs. autonomous system? How does taking the man out extend performance?
- What level of confidence is there in auto-coders?
- What different ways are there of measuring and reporting uncertainty?
- Performance characteristics to develop trust?
- It would be interesting to have an Autonomy Decathlon.

The final wrap-up concluded with thanks from the AFRL sponsors and a re-iteration of how the information would be used to feed the following workshops and ultimately input into an AFRL TEV&V strategy for autonomy. Kris previewed the storyboard that was created to tie the research portfolios to the AFRL Autonomy Strategy goal "Highly effective Human-Machine Teaming". This portrayal has been useful for that goal. As the ideas from this workshop are integrated with the ideas from the academia and industry, AFRL may want to develop something like this for T&E, V&V. The storyboard can be found in Appendix A4, "Highly effective Human-Machine Teaming" storyboard.

Academic Exploration Workshop – January 28-29, 2014

This section attempts to capture the discussion and outcome of the second of three workshops facilitated by the Wright Brothers Research Institute. This workshop included key individuals from Academia that have experience with creating / fielding autonomous systems. The Wright Brothers Institute did not limit the participants to just DoD nor to just AF. Additionally, the invitation was not intended to exclude other applications of autonomous systems not in the traditional robotic or vehicle space, autonomous cyber systems and medical systems were also welcome. The goal was to have as broad of a field as possible. The participants and their contact information is listed in Table 1 below

Wright Brothers Institute Facilitators

Cheryl Reed	WBI
Bart Barthelemy	WBI

AFRL Sponsors

Laura Humphrey	AFRL/RQQA
Kris Kearns	AFRL/RH
Jim Overholt	AFRL/RH
Matthew Clark	AFRL/RQQA

Industrial Participants

Behcet Acikmese	UTEXAS
Darryl Ahner	AFIT
Nick Armstrong-Crews	MIT
Dionisio de Niz	SEI/CMU
Georgios Fainekos	ASU
Karen Feigh	GATECH
Naira Hovakimyan	ILLINOIS
Lyle Long	PSU
Sandeep Neema	VANDERBILT
Bruce Preiss	WRIGHT STATE
Sanjai Rayadurgam	UMN
Rusty Roberts	GATECH
Rich Rowland	GATECH
Scott Stoller	STONYBROOK
Brian Stone	AFIT
Janos Sztipanovits	VANDERBILT
Lora Weiss	GATECH
Mick West	GATECH
David Woods	OSU
Enric Xargay	ILLINOIS
William Young	MIT

Workshop Agenda

3. DAY 1 - PROBLEM SPACE EXPLORATION

3.1. *ALIGN* – Introductions and Objectives

3.1.1. DOD Autonomy Intro – Kris Kearns, AFRL

3.1.2. TEV&V for AFRL – Matt Clark, AFRL

3.2. EXPLORE-PARTICIPANT INTROS – Role in TEV&V Biggest tech challenge

3.3. OPEN APERTURE – Discussions

3.3.1. What does certified mean?

3.3.2. What is an autonomous system?

***3.3.3.* What new innovations in software / hardware certification can be applied to Autonomous Systems?**

3.4. CONCEPT MAP DEVELOPMENT

4. DAY 2– CONVERGE ON CHALLENGES

4.1. ORIENT - Revisit Concept Map for new insights

4.2. DISTILL KEY ELEMENTS

4.2.1. Identify major technical thrusts from Concept map

4.2.2. Elaborate on/refine each technical thrust – what needs to be done to accomplish each

4.3. SEQUENCE CHALLENGES AND DEPENDENCIES

4.3.1. Determine any required sequencing of the various technical thrusts

4.4. WRAP UP

4.4.1. What is your biggest takeaway?

4.4.2. What would you ask / tell the Academic Workshop Participants?

Workshop Notes

3. DAY 1 - PROBLEM SPACE EXPLORATION

3.1. *ALIGN* – Introductions and Objectives

3.1.1. DOD Autonomy Intro – Kris Kearns, AFRL

The event started introducing the DoD Autonomy Technology Challenge Areas: Human and Agent Teaming, Scalable Teaming, Machine Reasoning and Intelligence, and ***Test and Evaluation, Verification and Validation***. Then identifying the AFRL Autonomy Goals: Deliver flexible autonomy systems with highly effective human-machine teaming, Create actively coordinated teams of multiple machines to achieve mission goals, Ensure operations in complex, contested environments, and ***Ensure safe and effective systems in unanticipated and dynamic environments***.

It was highlighted that the DoD and AFRL goals, although not the same, are similar. Specifically, the Test and Evaluation, Verification and Validation goal is directly attributed to the AFRL Goal of "ensuring safe and effective systems." The introduction demonstrated the AF and DoD need to understand and invest in new methods of assuring the safety and operability of up-and-coming autonomous systems. Additionally, Mrs. Kearns is responsible for coordinating an AFRL strategy for Autonomy. An overarching AFRL strategy document was signed as of Friday Oct 25th. The strategy now needs "meat on the bones" highlighting a portfolio that will accomplish the strategy goals. The human-machine teaming portfolio was created first and the TEV&V was determined to be the next portfolio to be tackled. This workshop is the first of three to aid in that goal.

3.1.2. TEV&V for AFRL – Matt Clark, AFRL

Mr. Clark continued the conversation by highlighting how the information from the three workshops would be used to provide the portfolio needed to accomplish the AFRL autonomy strategy. First, the industrial based workshop would identify the key certification challenges facing some of the DoD and other industries trying to enable further autonomous systems. The plan is to take the outcome of the meeting this week to feed to a similar academic workshop. The compiled information would then be distilled and coordinated with a broader AFRL Autonomy team. This team would consolidate and make recommendations for future TEV&V investments. Additionally, in parallel, the DoD Autonomy initiative is working to put together similar strategies. The inputs from the above three workshops will feed the greater DoD Autonomy effort, highlighting the investment opportunities and research needs.

3.2. EXPLORE-PARTICIPANT INTROS – Role in TEV&V Biggest tech challenge

During the section each participant was asked to provide some information about their background and one biggest challenge they face in TEV&V right now. These challenges were captured but not used until the end of the first day for the purposes of comparing the initial thoughts with the group challenges. These challenges seemed to converge to six overarching categories. Some of the challenges included:

- **Requirements, Models, and Design**
  - Defining requirements well, Validation of models, Unified framework for analysis
  - Define formal requirements for design and human/machine interaction
  - Awareness, limitations, aspects, chunks, and breakdown of Autonomy
  - Predictability of Failure
  - Defining performance metrics including human performance
  - Optimize design for high performance (non-conservative)
- **Human-Machine Interaction**
  - Multi-human operators/device
  - Autonomous & Semiautonomous human/machine interaction
  - Better communication/feedback w/ UAS
- **Modeling and Simulation, Testing**
  - Context for testing, Characterizing the environment
  - Tasks for testing (including human limits)
  - DOE w/ non deterministic/autonomous systems
  - M&S for predictability
  - Applied statistical techniques to emergent behavior

- Progressive, sequential testing w/ UAS
- Metrics/measurements against arguments

- **Runtime Assurance, Verification**
  - Understanding boundary conditions
  - Define and design safety envelopes
- **Highly Complex Interactive Autonomy**
  - How to deal with uncertainty
  - Understanding emergent behavior
  - Complexity/ resilience/ brittleness
  - Non-deterministic aspects
  - Scalability, Adaptability
- **Policy – User Impact**
  - Articulating what it brings to warfighter

3.3. OPEN APERTURE – Discussions

The next activity involved group discussions at each of the six tables asking three questions. The purpose of the exercise was to highlight the differences in how we define certification, autonomous, and new innovations in software / hardware certification can be applied to Autonomous Systems.

3.3.1. What does certified mean?

Some of the responses for what does certified mean:

- **<u>Third Party</u>** evidence based assessment that a system performs consistent to a specification or standard.
  - (NOT V&V)
- Certificate by competent authority that attests that the article presented passes the qualifying criteria. Some challenges to that are:
  - Autonomous system characteristics are different and break our normal certification process
  - The need for recertification
- Providing evidence of proof to a governing body that the system satisfies some properties that comply with previously established standards and can be **<u>trusted.</u>**
- Statement or assertion from an authoritative body that a **<u>person, object or system</u>** meets predetermined criteria and is validated to provide capability in a specified environment within certain constraints
- Certified system is one that an evaluation by a third party meets the performance, safety, and robustness requirement and every failure of which within is its prescribed operational envelope is a safe one.

3.3.2. What is an autonomous system?

A unique statement was made early in the morning about the caution in exploring the definition of Autonomy, Autonomous Systems, or levels of Autonomy. The concern was that many organizations have wasted a considerable amount of time debating the perfect definition autonomy rather than articulating the challenges needed to overcome the issues related with realizing it. Interestingly, the discussion on the definition of Autonomy did not vary as wildly as in the Industry workshop and was not as diverse as the discussion on what certification meant. Most groups agreed that Autonomy generally described a system with higher level of self-governing and less human interaction. As in the Industry workshop, Mrs. Kearns brought out the AFRL definition of Autonomy to guide the discussion:

"Automation: The system functions with no/little human operator involvement; however, the system performance is limited to the specific actions it has been designed to do.[9] Typically these are well-defined tasks that have predetermined responses, i.e. simple rule-based responses.

Autonomy: Systems which have a set of intelligence-based capabilities that allow it to respond to situations that were not pre-programmed or anticipated in the design (i.e. decision-based responses).

[9] Spacecraft Autonomy Technology: A Survey. Erwin, R. Scott and Paul Zetocha, AFRL/RV

Autonomous systems have a degree of self-government and self-directed behavior (with the human's proxy for decisions)."[10]

3.3.3. What new innovations in software / hardware certification can be applied to Autonomous Systems?

Based on the industry responses to the question "what does it mean to certify an autonomous system" we decided to change the question for the academic crowd. Unfortunately, the intent of the question was not articulated well by Mr. Clark and the audience interpreted and answered the question differently than expected. Originally, the hope was to highlight new and innovative research that might help with the V&V of Autonomous systems. However, the responses indicated additional needs / challenges. A full list of these challenges can be found on page 5 of appendix A1 attached within this document.

3.4. CONCEPT MAP DEVELOPMENT

The second half of the first day focused on taking the assumptions about certification and autonomy and using them to produce a concept map. Each table was directed to come up with the 5 top challenges to the TEV&V of autonomous systems independently. Once completed, each table was directed to consolidate "like" or common challenges. Finally, each table took turns placing their consolidated challenges up on the white board, again grouping them based on similarity. These became the overarching categories that objectives and technical challenges were defined on day two. Under each category, a list of similar challenges was identified. These challenges can be found on the following pages in appendix A1. The miscellaneous category was dissolved on the second day.

Challenge "Ideation"	Page
Human automation interaction	7
Requirement generation	10
System assurance methodology	13
Standards & architecture	16
Capabilities & limitations	19
Emergence	22
Learning & memory	25
Complexity	28
Security	31
Teaming of multiple entities	34
Misc	37

4. DAY 2– CONVERGE ON CHALLENGES

4.1. ORIENT - Revisit Concept Map for new insights

The academic group did highlight a concept similar to the industry group's "Training" and then "Licensing" a human vs. Certifying a machine or system. They called it "Learning & Memory." This concept did not resonate as deeply with the academic group but several participants identified a need to dramatically change how certification is accomplished now.

The most dramatic realization was increased buy-in to the Wright Brother's Research Institute collaboration process. Several individuals were skeptical that the workshop would produce any meaningful results. However, the second day, the same individuals felt that the categories and specifically the crosscutting interdependencies between the categories were particularly interesting.

[10] Air Force Research Laboratory Autonomy Science and Technology Strategy: Maj Gen Masiello, Oct 2013

4.2. DISTILL KEY ELEMENTS

4.2.1. Identify major technical thrusts from Concept map

4.2.2. Elaborate on/refine each technical thrust – what needs to be done to accomplish each

The next exercise took the bulk of the morning and into the afternoon. The participants were directed to take each top level challenge identified previously and perform a mini GOTCHA exercise. The top level challenges were considered the Goal. Each team then identified the objective and technical challenges to be addressed to achieve the goal. The below list highlights the objectives of each challenge identified. The Technical Challenges can be found in appendix A1.

Human automation interaction(HAI)

Objective:

- To create a T&EV&V system that achieve HAI that mirrors/ improves upon the best human-human teams
- Creation of a transparent and traceable, predictable human-machine interface for the autonomous system

Requirement generation

Objective:

- Requirements should have the following properties:
  - Consistent
  - Testable (objectively, measurable, formally)
  - Traceability to goals and through levels of abstraction
  - Process and outcomes
  - Must account for nominal and off-nominal situations
  - Context sensitive
- Requirements currently specify what system must do, but should specify all aspects of the system
- To generate comprehensive and objectively testable (evaluate-able) requirements for a joint human-machine systems at multiple levels in such a way as to systematically build an “evidence case” that through this joint system meets performance objectives while preserving some level of safety or fail safe modes in off normal situations

System assurance methodology

Objective:

- Develop methods to assure that systems meet evolving requirements, operational needs, and certification criteria, and continue to meet them throughout the system lifecycle, including development and operational phases.

Standards & architecture

Objective:

- Specify interfaces, protocols, and data formats in support of T&EV&V across stakeholders
  - Who does this? Who enforces?
  - Including “Joint Test Action Group (JTAG)” like interfaces for instrumentation
- Standardize functional layers for autonomous systems (e.g. OSI)
  - Specify and enforce design practices for testability and certification (component-based design, separation of data and algorithms)

Capabilities & limitations

Objective:

- Continuously identify capabilities and limitations as a function of the mission (context, objectives, constraints)

Emergence

Objective:

- Multiple interacting entities at different levels of analysis exhibit emergent properties.

- What T&EV&V methods characterize the properties of a system of multiple interacting entities and if/how those properties contribute to mission success or risk.
    - Properties exist interacting entities

Learning & memory

Objective:

- Provide evidence that the system adapts and improves behavior as it accumulates operational data
- Test if it continues to meet certification requirements
- Quantify the potential bounds of the learned behavior

Complexity

Objective:

- Autonomous systems are complex because of multitude of entities and cross cutting interactions
- Objective is to innovate compositional methods for formal and empirical T&EV&V with a goal of discovering and managing hidden side effects

Security

Objective:

- Interaction with learning opens up a new threat vector (force a bad learning path)
- Determining effective red teaming approaches for autonomous systems

Teaming of multiple entities

Objective:

- Create a T&EV&V process that exposes and manages the mission actions/interactions of multiple entities
    - Team structure
    - Performance in a particular mission
    - Drive down exponential cost and complexity
    - LVC approach?

4.3. SEQUENCE CHALLENGES AND DEPENDENCIES

4.3.1. Determine any required sequencing of the various technical thrusts

The next ½ of the day focused on prioritizing the major challenges to verifying and validating autonomous systems and technical challenges needed to achieve them. Figure 2 shows the compiled results of what major challenges the academic group recommended to address in the near, mid, and far term.

Challenge	Green	Yellow	Red
Requirement generation	13	3	2
Standards & architecture	12	4	1
Human automation interaction	9	5	1
System assurance methodology	6	5	4
Learning & memory	5	7	7
Security	4	8	4
Teaming of multiple entities	3	4	12
Capabilities & limitations	2	12	5
Complexity	2	5	10
Emergence	0	4	11

Figure 2: Priority and Score for Each Technical Challenge
(Green = Near Term, Yellow = Mid Term, Red = Far Term)

The academic group also highlighted what they referred to as “cross-cutting” technical challenges. This referred to technical challenges that needed to be addressed to make progress on several areas. Cross cutting technical challenges were marked with a C in appendix A1.

Additionally, the group was given 5 blue dots to emulate 5 “dollars” they would spend on any particular technical challenge. These are marked in appendix A1 with a number and followed by the letter B. The number indicates how many “dollars” were invested in that technical challenge.

WRAP UP

4.3.2. What is your biggest takeaway?

Some of the biggest takeaways are highlighted below. More detail can be found in appendix A1 under Final Thoughts.

- As automated systems become more and more human like, the certification process is going to become more like the process used on humans.
- Suggestion: do an idea map and map out the interconnections between the different topics.
- Science and basic research are not currently on track to tackle T&EV&V, there needs to be a shift in thinking and a whole new way to tackle the issue.
- There can be a language barrier in the discussion. A common language is necessary between academia and industry.
- Only way to speed up development, is if a major entity puts major money behind it.
- What capabilities are lacking? Build capabilities before architecture.
- Industry needs something tangible to test.
- A giant challenge problem is needed to push things forward. (e.g. DARPA)
- Would need to be open source
- There is a need to change public opinion on what a robot is and can do.
- Can you predict what/when/why the system will fail?
- The things that came up on top are already being worked on in other domains. How can things be applied differently? What synergies exist?
- Academics don’t always have the ability to scale developments to the necessary level to truly test the idea.
- Developing a testing framework is very difficult. What assumptions are you making, and what are you assuming away? What gaps exist in the framework?
- While the focus of the workshop was T&EV&V, the focus and problem is the upfront design process and the system itself.
- Modularization needs to be pushed more and further. How will things connect and work together?
- There is a need for experimental platforms.
- The problems are not unique to autonomous systems.
- There was a big focus on the engineering issues for autonomous systems. Can’t put the cart before the horse.

4.3.3. What would you ask / tell the Industry Workshop Participants?

When asked the questions that industry had for them, the academic group had the following return questions:

- Does industry have any capabilities developed?
- How do you want to move forward on this? Who is funding?
- How would you like to partner with academics? What is the process by which to do so? What would be the key to enabling technology to transition?
- What does a program look like that would reap benefits from both Industry and Academia?
- How can we work together to reframe/tackle the issue?
- Are you (Industry) a better group to design/fund a challenge or to test some of these ideas (decathlon)
- What is necessary to convince the FAA?
- Do you know what you want/need?
- Can you share some generalized architecture without giving up valuable proprietary developments?
- Are you interested in modular architectures, how are you moving towards modular technology?
- How do academics and industry engage in a dialogue? How can the dialogue be beneficial for both parties?
- Could you provide more resolution on the regulatory issues which are preventing the introduction of developed technologies?

The final wrap-up concluded with thanks from the AFRL sponsors and a re-iteration of how the information would be used to feed the following workshops and ultimately input into an AFRL TEV&V strategy for autonomy. Matt presented the summary of the industrial workshop, attached as appendix A2. Kris previewed the storyboard that was created to tie the research portfolios to the AFRL Autonomy Strategy goal "Highly effective Human-Machine Teaming". This portrayal has been useful for that goal. As the ideas from this workshop are integrated with the ideas from the academia and industry, AFRL may want to develop something like this for T&E, V&V. The storyboard can be found in Appendix A4, "Highly effective Human-Machine Teaming" storyboard.

Government Exploration Workshop – February 25-26, 2014

This section attempts to capture the discussion and outcome of the third of three workshops facilitated by the Wright Brothers Research Institute. This workshop included key government individuals from the Air Force Research Laboratory and the Air Force Test Center that have experience with creating / fielding autonomous systems. The Wright Brothers Institute made an effort to include participants from other applications of autonomous systems not in the traditional robotic or vehicle space. For example, autonomous cyber systems and medical systems were also welcome. The goal was to have as broad of a field as possible and to have representation from each technical directorate. The participants and their contact information is listed in the tables below

Wright Brothers Institute Facilitators

Cheryl Reed	WBI
Bart Barthelemy	WBI

AFRL Sponsors

Kerianne Gross	AFRL/RQQD
Kris Kearns	AFRL/RH
Jim Overholt	AFRL/RH
Matthew Clark	AFRL/RQQA

Government Participants

Brian Abbe	AFRL/RI
Scott Douglass	AFRL/RH
Richard Erwin	AFRL/RV
Kevin Gluck	AFRL/RH
William Gray	TPS/CP
Bill Koenig	AFRL/RY
Raj Malhotra	AFRL/RY
Richard Metzger	AFRL/RI
Joseph Nichols	AFTC/CZ
Andy Rice	AFRL/RY
Corey Schumacher	AFRL/RQ
Robert Smith	Robert Smith
Michael Talbert	AFRL/RY
Daniel Thompson	AFRL/RQ
Tony Thompson	AFRL/RW
Ryan Turner	AFRL/RI
Lok Yan	AFRL/RI
Paul Zetocha	AFRL/RV

Workshop Agenda

5. DAY 1 - PROBLEM SPACE EXPLORATION

5.1. *ALIGN* – Introductions and Objectives

5.1.1. DOD Autonomy Intro – Kris Kearns, AFRL

5.1.2. TEV&V for AFRL – Matt Clark, AFRL

5.2. EXPLORE-PARTICIPANT INTROS – Role in TEV&V Biggest tech challenge

5.3. OPEN APERTURE – Discussions

5.3.1. What does certified mean?

***5.3.2.* What does it mean to certify an Autonomous System**

5.4. EXPLORE/DIVERGE – Concept map the challenges

5.5. DISTILL KEY ELEMENTS

5.5.1. Identify major technical thrusts from Concept map

5.5.2. Elaborate on/refine each technical thrust – what needs to be done to accomplish each

5.6. SEQUENCE CHALLENGES AND DEPENDENCIES

5.6.1. Determine any required sequencing of the various technical thrusts

6. DAY 2– CONVERGE ON CHALLENGES

6.1. Overview of previous T&E V&V For Autonomy forums – Matt Clark

6.2. Brief of AFRL Strategy and the Human Machine Teaming story board – Jim Overholt

6.3. 2030 Vision - what we need T&E V&V of autonomy to look like

6.4. Synthesize objectives into key technical trusts

6.5. WRAP UP

6.5.1. What is your biggest takeaway?

6.5.2. What would you ask / tell the Academic Workshop Participants?

Workshop Notes

5. *DAY 1 - PROBLEM SPACE EXPLORATION*

5.1. *ALIGN* – Introductions and Objectives

The introduction and objectives followed the same format as the last two workshops. Ms. Kearns introduced the AFRL Autonomy Strategy and Mr. Clark introduced the V&V of Autonomy area. It was presented that the schedule for the government workshop would be slightly different than the last two. While the Industry and Academic groups took two days to generate a set of Objectives and Technical Challenges in TEV&V, the government was going to generate these artifacts in one day. The second day of the workshop would be focused on aggregating the data from all three workshops, striving to come up with the top 4-6 technical objectives. A final meeting with the Autonomy TEV&V meeting owners will process the data from all three workshops to generate a draft storyboard and report.

5.2. EXPLORE-PARTICIPANT INTROS – Role in TEV&V Biggest tech challenge

During the section each participant was asked to provide some information about their background and one biggest challenge they face in TEV&V right now. Some of the challenges included:

- Emergent behavior - non-determinism (predictability and characterization)
- Certification/ accreditation of Autonomous System
- Unknown, actively hostile environment of military Autonomous System
- Buy-in from T&EV&V community
- Validly decompose T&EV&V problem
- Prediction, characterization, control of Autonomous System
- Dealing with non-determinism of Autonomous System
- Prediction of systems that use knowledge
- Integration of behavior descriptions at different levels of abstraction
- Testing of self-governing/self-behaving systems challenges
- Difficulty of objective assessment of self-governing algorithms
- Composability, non-determinism aspects of Autonomous Systems with testing
- Defining a process to show Autonomous System doesn't do what you don't want it to do
- Valid modeling and simulation of human component of Autonomous System
- Regulatory problems and barriers
- Moving from lab to field introduces new issues and unknown threats
- V&V processes not set up for non-deterministic systems

5.3. OPEN APERTURE – Discussions

The next exercise asked two questions, what does certify mean and what does it mean to certify an autonomous system? Due to the compressed nature of the Government workshop, the definitions of certification and certification of autonomous systems bled into the concept map challenges. The purpose of this exercise was primarily focused on getting discussions started around these two questions. Unlike the first two workshops, the different groups in the workshop did not deviate that much from a common definition of certification as the acknowledgement from an authoritative body that a particular system complies with a defined standard.

5.4. *EXPLORE/DIVERGE – Concept map the challenges*

The next section identified the common challenges. These challenges formed the objectives for the afternoon. A detailed list of the challenges identified can be found in appendix A7. The summarized challenges:

Policy & Standards

Objective:

- Established communication plan between designers and policy makers influence policy changes
    - Training vs. test

Certification of Unpredictable Systems

Objective:

- Enable certification of highly complex systems with large, unpredictable state spaces with:
    - Incomplete specifications, Adaptive performance, Nonlinearities, Uncertainties, Time variant behavior

Runtime Assurance & Monitoring

Objective:

- Instrument autonomous systems with runtime monitoring/tie in at design time
- Monitor a total system behavior for undesirable states at run time and if necessary, activate alternative controller (i.e. man, autonomous system)

Cultural Acceptance

Objective:

- Develop best practices for TEVV of autonomous systems for technology transition and to ensure public trust
- Develop test and verification specifically to address human safety, privacy, morality
- Develop test VV practices to establish understanding, predictability, and degree of transparency

Human-Autonomy Interaction

Objective:

- Assess the extent to which the interaction of the human and autonomous components achieves the system requirements

Formal Methods

Objective:

- Reduce reliance on exhaustive test/stimulation by adopting proofs as evidence for certification
- Move V&V earlier in the design process by precisely specifying architectural requirements from the beginning

Efficient Testing

Objective:

- Create cost-effective testing via tools and techniques that identify and correct problems earlier in the systems engineering cycle
    - More accurate modeling and simulation
    - Analytical proof
    - Adaptive production of test
    - Virtual prototypes/test

System Compose-ability & Recertification

Objective:

- Efficient re-certification
- Reduce cost and effort associated with V&V of new system of system interactions

Moral Software Constructs encoded in future "evolving" or self-programming agents

Objective:

- Ensure benign intent of eventual super intelligent, rapidly-evolving AI systems
- Utility function engineering and coherent extrapolated volition (AI drives)
- safeguards (e.g. RTA), firewalls, security, and emergent behavior protections
- real-time introspection and verification of learning, self-modifying software

5.5. DISTILL KEY ELEMENTS

As with the industry and academic workshops, the participants were directed to take each top level challenge identified previously and perform a mini GOTCHA exercise. The top level challenges were considered the Goal. Each team then identified the objective and technical challenges to be addressed to achieve the goal. A detailed list of the challenges identified can be found in appendix A7.

5.6. SEQUENCE CHALLENGES AND DEPENDENCIES

5.6.1. Determine any required sequencing of the various technical thrusts

The next ½ of the day focused on prioritizing the major challenges to verifying and validating autonomous systems and technical challenges needed to achieve them. The Figure below shows the compiled results of what major challenges the government group group recommended to address in the near, mid, and far term.

Challenge	Near	Mid	Far
Certification of Unpredictable Systems	13	2	0
Formal Methods	8	4	3
Efficient Testing	7	5	1
Runtime Assurance & Monitoring	6	5	4
System Compose-ability & Recertification	6	7	1
Human-Autonomy Interaction	4	9	4
Unknowable Environment	3	9	6
Policy & Standards	1	5	12
Cultural Acceptance	1	3	10

Priority and Score for Each Technical Challenge
(Green = Near Term, Yellow = Mid Term, Red = Far Term)

6. DAY 2– CONVERGE ON CHALLENGES

6.1. Overview of previous T&E V&V For Autonomy forums – Matt Clark

On day 2 the government group was asked to synthesize the work done by all three working groups. To begin the meeting, Mr. Clark presented an overview briefing of the previous two workshop outcomes and an initial aggregate view of common themes that seemed to be emerging between all three groups. Appendix A8 contains the technical thrusts, objectives, and technical challenges for all three workshops in one briefing. Appendix A9 contains the overview brief presented by Mr. Clark. The figure below attempts to illustrate emerging themes between the three workshops. At the beginning of Day 2, it was clear that eight unique thrusts began to emerge. Additionally, there was some debate as to what constitutes a technical thrust vs. a challenge or an enduring problem. For example, grouping four highlights V&V of Autonomy challenges like uncertainty, emergent behavior, complexity, etc. The Government team felt that these could be considered enduring problems to overcome.

Industry				Academic				Government			
Challenge	Near	Mid	Far	Challenge	Near	Mid	Far	Challenge	Near	Mid	Far
Modeling and Simulation	15	2	2					Efficient Testing	7	5	1
Test	11	6									
Defacto Standards Licensing vs Certification	14	2	1	Standards & architecture	12	4	1	Policy & Standards	1	5	12
Trust	4	10	3					Cultural Acceptance	1	3	10
Tool Verification	3	9	4				2	Formal Methods	8	4	3
Requirements	14	5	11	Requirement generation	13	3					
Synthesis	1	4	15								
				Learning & memory	5	7	7	Certification of Unpredictable Systems	13	2	0
Uncertainty	6	6	5	Complexity	2	5	10	Unknowable Environment	3	9	6
Emegent Behavior	1	3	18	Emergence	0	4	11				
Human/Machine Interaction	2	12	4	Human automation interaction	9	5	1	Human-Autonomy Interaction	4	9	4
Runtime Verification	2	11	5	Capabilities & limitations	2	12	5	Runtime Assurance & Monitoring	6	5	4
System of Systems V&V	3	8	7	System assurance methodology	6	5	4	System Compose-ability & Recertification	6	7	1
Security	4	9	8	Security	4	8	4				
				Teaming of multiple entities	3	4	12				

Priority and Score for Each Technical Challenge by workshop. Grouped into 8 categories (Green = Near Term, Yellow = Mid Term, Red = Far Term)

6.2. Brief of AFRL Strategy and the Human Machine Teaming story board – Jim Overholt

Dr. Overholt re-iterated how the information would be used to feed the following workshops and ultimately input into an AFRL TEV&V strategy for autonomy. He previewed the storyboard that was created to tie the research portfolios to the AFRL Autonomy Strategy goal "Highly effective Human-Machine Teaming". This portrayal has been useful for that goal. As stated earlier, the storyboard can be found in Appendix A4, "Highly effective Human-Machine Teaming".

6.3. 2030 Vision - what we need T&E V&V of autonomy to look like

This exercise concentrated on the technical thrusts from each workshop, trying to pull together a common vision. The workshop was divided into four groups. Each group put together their vision of 2030 for TEV&V of Autonomy. Some highlights are shown in the list below. The full contribution by each government group can be found in appendix A7 starting with slide 25.

2030 Vision:

- Fast certification and recertification, including of system changes
- Composability of certification for system of system
- Self-testing systems-monitor and expand own performance envelope
- Integrated safeguards and active and continuous V&V (embedded in instrumentation)
- Autonomy that can be "trained" and certified like a Pilot
- Reusable cases for the assurance of Autonomous systems and components
- Common, objective, formal semantics to define:
    - Test, proof, runtime constraints
    - Risk (probability based)
- Analyze acceptable levels of risk based on new missions, system compositions
- Common tool suite that includes domain specific languages
- Multi-agent collaboration technologies, including:
    - Secure trusted information sharing
    - Formal methods for prediction of emergent behavior
    - RTA of multi-agent/swarm systems
- International agreements/conventions on autonomy

6.4. Synthesize objectives into key technical trusts

Finally, the government team was asked to take the data from all three workshops and come up with 3 to 6 common technical thrusts that align with the vision statements in 6.3. The group was able to narrow down the technical objectives to 14, listed below.

- Run Time Assurance
- V&V in Early Design & Specification
- Active & Continuing V&V
- Human/Autonomy Interaction
- Dynamic Modeling & Analysis of Complexity
- Formal Models
- Policy
- T&EV&V of Multi-system Interaction
- T&EV&V of Learning Systems
- Validating Safeguards
- T&EV&V of Impact from Unknowable Environments
- Efficient Test Tools & Procedures
- Transparency
- Integrative Design, Modeling & Testing

Further discussions highlighted that some of the technical thrusts from the three workshops were cross cutting. Rather than try to consolidate down to 6 thrust areas, the team highlighted how the technical thrusts from all three workshops contributed to the consolidated list. Appendix A10 documents the interplay between the above 14 consolidated technical thrusts and the technical thrusts generated from each workshop. Appendix A8 contains the technical thrusts, objectives, and technical challenges for all three workshops in one briefing for reference. The final strategy can be found (graphically) in Appendix 11, and described in the Executive Summary.